

CANONICAL AND QUADRATIC FORMS

4.1 Jordan Canonical Form: Basic Properties

Recall that an $n \times n$ (complex) matrix A is diagonalizable if and only if A has a full set of n linearly independent eigenvectors, or equivalently, the dimension of each eigenspace $E(\lambda) = \mathcal{N}(\lambda I - A)$ is equal to the multiplicity of the eigenvalue λ . Hence, if $\lambda_1, \dots, \lambda_t$ are distinct eigenvalues of A with multiplicities $m_{\lambda_1}, \dots, m_{\lambda_t}$, respectively, then

$$\dim E(\lambda_1) + \dots + \dim E(\lambda_t) = m_{\lambda_1} + \dots + m_{\lambda_t} = n$$

and

$$\mathbb{C}^n = E(\lambda_1) \oplus \dots \oplus E(\lambda_t).$$

On the other hand, a matrix A is *not* diagonalizable if and only if A has an eigenvalue λ with multiplicity $m_\lambda > 1$ such that

$$1 \leq \dim E(\lambda) < m_\lambda,$$

so that the number of linearly independent eigenvectors belonging to λ must be less than m_λ .

However, even if a matrix A is not diagonalizable, one may try to find a matrix similar to A which has as many zero entries as possible except diagonals. Schur's lemma says that any square matrix is similar to an upper triangular matrix. In fact, any square matrix A can be similar to a matrix much "closer" to a diagonal matrix, called a *Jordan canonical form*. Its diagonal entries are the eigenvalues of A , the entry just above each diagonal entry is 0 or 1, and all other entries are 0. In this case, the columns of a basis-change matrix Q are something like eigenvectors, but not the same in general. They are called *generalized eigenvectors*.

Theorem 4.1.1. *For any square matrix A , A is similar to a matrix J of the following form,*

called the **Jordan canonical form** of A or a **Jordan canonical matrix**,

$$J = Q^{-1}AQ = \begin{bmatrix} J_1 & & \mathbf{0} \\ & J_2 & \\ \mathbf{0} & & \ddots \\ & & & J_s \end{bmatrix},$$

in which

- (1) s is the number of linearly independent eigenvectors of A , and
- (2) each J_i is an upper triangular matrix of the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \mathbf{0} \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \mathbf{0} & & & \lambda_i \end{bmatrix},$$

where λ_i is a single eigenvalue of J_i with only one linearly independent associated eigenvector. Such J_i is called a **Jordan block** belonging to the eigenvalue λ_i .

Proof. Beyond the scope of our text and syllabus. □

Note that if an $n \times n$ matrix A has a full set of n linearly independent eigenvectors (that is $s = n$), then there have to be n Jordan blocks so that each Jordan block is just a 1×1 matrix, and an eigenvalue λ appears as many times as its multiplicity. In this case, the Jordan canonical form J of A is nothing but the diagonal matrix with eigenvalues on the diagonal and a basis-change matrix Q is defined by taking the n linearly independent eigenvectors as its columns. Thus, a diagonal matrix is a particular case of the Jordan canonical form.

- Remark 4.1.2.** (1) For a given Jordan canonical form J of A , one can get another one by permuting the Jordan blocks of J , and this new one is also similar to A . In fact, we shall see later that, any two Jordan canonical forms of A cannot be similar except this possibility. Thus, any (complex) square matrix A is similar to only one Jordan canonical matrix up to permutations of the Jordan blocks. In this sense, it is called *the* Jordan canonical form of a matrix A .
- (2) As an alternative way to define the Jordan canonical form of A , one can take the transpose J^T of the Jordan canonical matrix J given in the above theorem. In this case, each Jordan block becomes a lower triangular matrix with a single eigenvalue. However, this definition does not make any difference.

The following example shows which matrices A have the Jordan canonical form J and how the s linearly independent eigenvectors of A (or J) correspond to the Jordan blocks in J .

Example 4.1.3 (Each Jordan block to each eigenvector). Let J be a Jordan canonical matrix of the form:

$$J = \begin{bmatrix} \begin{bmatrix} 6 & 1 \\ 0 & 6 \end{bmatrix} & & \\ & \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} & \\ & & [2] \end{bmatrix} = \begin{bmatrix} J_1 & & \\ & J_2 & \\ & & J_3 \end{bmatrix}.$$

Find the number of linearly independent eigenvectors of J and determine all matrices whose Jordan canonical forms are J .

Solution: Since J is a triangular matrix, the eigenvalues of J are the diagonal entries 6 and 2 with multiplicities 2 and 3, respectively. The eigenspace $E(6)$ has a single linearly independent eigenvector because by the Rank Theorem (Theorem 1.6.5),

$$\dim E(6) = \dim \mathcal{N}(J - 6I) = 5 - \text{rank}(J - 6I) = 5 - 4 = 1.$$

To find the eigenspace $E(6)$ of the matrix, the matrix equation $(J - 6I)\mathbf{x} = \mathbf{0}$ yields

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

This gives $x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0$. Letting the free parameter $x_1 = t$ ($t \in \mathbb{R}$), we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} t \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = t\mathbf{e}_1.$$

Thus, the eigenspace is one-dimensional and is given by

$$E(6) = \{t\mathbf{e}_1 \mid t \in \mathbb{R}\} \Rightarrow \dim E(6) = 1.$$

This single basis vector $\mathbf{e}_1 = (1, 0, 0, 0, 0)$ corresponds directly to the single Jordan block J_1 associated with the eigenvalue $\lambda = 6$.

Similarly, we can see that the eigenspace $E(2)$ has two linearly independent eigenvectors $\mathbf{e}_3$ and $\mathbf{e}_5$ with

$$\dim E(2) = 5 - \text{rank}(J - 2I) = 5 - 3 = 2,$$

and $\lambda = 2$ appears in two blocks J_2 and J_3 .

Hence, if a matrix A is similar to J , then A is a 5×5 matrix whose eigenvalues are 6 and 2 with multiplicity 2 and 3 respectively, but there is only one linearly independent eigenvector belonging to 6, (i.e., $\dim E(6) = 1$) and there are only two linearly independent eigenvectors belonging to 2, (i.e., $\dim E(2) = 2$).

In general, if a matrix A is similar to a Jordan canonical matrix J , then both matrices have the same eigenvalues of the same multiplicities and $\dim \mathcal{N}(\lambda I - A) = \dim \mathcal{N}(\lambda I - J)$ for each eigenvalue λ . $\square$

Exercise 4.1.4. Show that for any Jordan block J , the order of J is the smallest positive integer k such that $(J - \lambda I)^k = \mathbf{0}$, where λ is an eigenvalue of J .

An eigenvalue λ may appear in several blocks. In fact, the number of Jordan blocks belonging to an eigenvalue λ is equal to the number of linearly independent eigenvectors of A belonging to λ , which is the dimension of the eigenspace $E(\lambda) = \mathcal{N}(\lambda I - A)$. Moreover, the sum of the orders of all Jordan blocks belonging to an eigenvalue λ is equal to the multiplicity m_λ of λ .

The following example shows all possible cases of Jordan blocks belonging to an eigenvalue λ when its multiplicity m_λ is fixed.

Example 4.1.5 (*Classifying Jordan canonical matrices having a single eigenvalue*). Classify all possible Jordan canonical forms of a 5×5 matrix A that has a single eigenvalue λ of multiplicity 5 (up to permutations of the Jordan blocks).

Solution: There are **seven** possible Jordan canonical forms as follows.

- (1) **Suppose A has only one linearly independent eigenvector belonging to λ .** Then the Jordan canonical form of A is of the form

$$J^{(1)} = Q^{-1}AQ = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix},$$

which consists of only one Jordan block with eigenvalue λ on the diagonal. And, both A and $J^{(1)}$ have only one linearly independent eigenvector belonging to λ . (Note that $\text{rank}(J^{(1)} - \lambda I) = 4$.)

- (2) **Suppose it has two linearly independent eigenvectors belonging to λ .** Then the Jordan canonical form of A is either one of the forms

$$J^{(2)} = \begin{bmatrix} \lambda & 1 & & & \\ 0 & \lambda & & & \\ & & \lambda & 1 & 0 \\ & & 0 & \lambda & 1 \\ & & 0 & 0 & \lambda \end{bmatrix} \quad \text{or} \quad J^{(3)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & 1 & 0 & 0 \\ & 0 & \lambda & 1 & 0 \\ & 0 & 0 & \lambda & 1 \\ & 0 & 0 & 0 & \lambda \end{bmatrix},$$

each of which consists of two Jordan blocks belonging to the eigenvalue λ . Note that $J^{(2)}$ has two linearly independent eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_3$, while $J^{(3)}$ has $\mathbf{e}_1$ and $\mathbf{e}_2$. These two matrices $J^{(2)}$ and $J^{(3)}$ cannot be similar, because $(J^{(2)} - \lambda I)^3 = \mathbf{0}$, but $(J^{(3)} - \lambda I)^3 \neq \mathbf{0}$. (can be justified by a direct computation).

- (3) **Suppose it has three linearly independent eigenvectors belonging to λ .** Then the Jordan canonical form of A is either one of the forms

$$J^{(4)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & 1 & & \\ & 0 & \lambda & & \\ & & & \lambda & 1 \\ & & & 0 & \lambda \end{bmatrix} \quad \text{or} \quad J^{(5)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & 1 & 0 \\ & & 0 & \lambda & 1 \\ & & 0 & 0 & \lambda \end{bmatrix},$$

each of which consists of three Jordan blocks belonging to the eigenvalue λ . Note that $J^{(4)}$ has three linearly independent eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_4$, while $J^{(5)}$ has $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. These two matrices $J^{(4)}$ and $J^{(5)}$ are not similar, because $(J^{(4)} - \lambda I)^2 = \mathbf{0}$, but $(J^{(5)} - \lambda I)^2 \neq \mathbf{0}$.

- (4) **Suppose it has four linearly independent eigenvectors belonging to λ .** Then the Jordan canonical form of A is of the form

$$J^{(6)} = \begin{bmatrix} \lambda & & & & \\ & \lambda & & & \\ & & \lambda & & \\ & & & \lambda & 1 \\ & & & 0 & \lambda \end{bmatrix},$$

which consists of four Jordan blocks with eigenvalue λ .

(5) **Suppose it has five linearly independent eigenvectors belonging to λ .** Then the Jordan canonical form of A is the diagonal matrix $J^{(7)}$ with diagonal entries λ .

$$J^{(7)} = \begin{bmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}.$$

Note that all of these seven possible Jordan canonical matrices have the same trace, determinant and the same characteristic polynomial, but no two of them are not similar to each other. $\square$

As shown in the case (2) (also in (3)) of the above example, two Jordan canonical matrices $J^{(2)}$ and $J^{(3)}$ have the same eigenvalue of the same multiplicity and they also have the same number of linearly independent eigenvectors, but they are not similar. The problem of choosing one of the two possible Jordan canonical forms which is similar to the given matrix A depends on the sequence of $\text{rank}(A - \lambda I)^\ell$ for $\ell = 1, 2, \dots, n$.

The next example illustrates how to determine the orders of the Jordan blocks belonging to the same eigenvalue λ .

Example 4.1.6 (*Determine Jordan blocks belonging to the same eigenvalue*). Let J be a Jordan canonical matrix with a single eigenvalue λ :

$$J = \begin{bmatrix} J_1 & & & \mathbf{0} \\ & J_2 & & \\ & & J_3 & \\ \mathbf{0} & & & J_4 \end{bmatrix} = \begin{bmatrix} \lambda & 1 & 0 & & & \\ 0 & \lambda & 1 & & & \\ 0 & 0 & \lambda & & & \\ & & & \lambda & 1 & \\ & & & 0 & \lambda & \\ & & & & & \lambda & 1 \\ & & & & & 0 & \lambda \\ & & & & & & & [\lambda] \end{bmatrix}.$$

Then,

$$(J - \lambda I)^2 = \begin{bmatrix} 0 & 0 & 1 & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \\ & & & 0 & 0 & \\ & & & 0 & 0 & \\ & & & & & 0 & 0 \\ & & & & & 0 & 0 \\ & & & & & & [0] \end{bmatrix} \quad \text{and} \quad (J - \lambda I)^3 = \mathbf{0}. \quad \square$$

Thus, one can get $\text{rank}(J - \lambda I) = 4$, $\text{rank}(J - \lambda I)^2 = 1$ and $\text{rank}(J - \lambda I)^3 = 0$. This sequence of ranks, $\text{rank}(J - \lambda I)^\ell$ for $\ell = 1, 2, 3$, determines completely the orders of blocks of J . In fact, one can notice that

- (i) the fact that $(J - \lambda I)^3 = \mathbf{0}$ but $(J - \lambda I)^2 \neq \mathbf{0}$ implies that the largest block has order 3,
- (ii) $\text{rank}(J - \lambda I)^2 = 1$ is equal to the number of blocks of order 3,
- (iii) $\text{rank}(J - \lambda I) = 4$ is equal to twice the number of blocks of order 3 plus the number of blocks of order 2, so there are two of them,

(iv) the number of blocks of order 1 is $8 - (2 \times 2) - (3 \times 1) = 1$.

Example 4.1.7 (*Jordan canonical form of a triangular matrix*). Find the Jordan canonical form J of the matrix

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}.$$

Solution: Since A is triangular, the eigenvalues of A are the diagonal entries $\lambda_1 = \lambda_2 = 2$, $\lambda_3 = 3$. Hence, there are two possibilities of the Jordan canonical form of A :

$$J^{(1)} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{or} \quad J^{(2)} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

But, one can see that $\text{rank}(A - 2I) = 2$. It implies that $\text{rank}(J - 2I) = 2$ and so the Jordan canonical form of A must be $J^{(2)}$. $\square$

Example 4.1.8 (*Jordan canonical form of a companion matrix*). Find the Jordan canonical form J of the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 4 & -6 & 4 \end{bmatrix}.$$

Solution:

$$\begin{aligned} \det(\lambda I - A) &= \begin{vmatrix} \lambda & -1 & 0 & 0 \\ 0 & \lambda & -1 & 0 \\ 0 & 0 & \lambda & -1 \\ 1 & -4 & 6 & \lambda - 4 \end{vmatrix} \\ &= \lambda \cdot M_{11} - 0 \cdot M_{21} + 0 \cdot M_{31} - 1 \cdot M_{41}, \end{aligned}$$

where M_{11} and M_{41} are the respective minors:

$$M_{11} = \begin{vmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -4 & 6 & \lambda - 4 \end{vmatrix}, \quad M_{41} = \begin{vmatrix} -1 & 0 & 0 \\ \lambda & -1 & 0 \\ 0 & \lambda & -1 \end{vmatrix}.$$

Note that $M_{41} = (-1) \cdot (-1) \cdot (-1) = -1$ and

$$\begin{aligned} M_{11} &= \lambda \begin{vmatrix} \lambda & -1 \\ 6 & \lambda - 4 \end{vmatrix} - 4 \begin{vmatrix} -1 & 0 \\ \lambda & -1 \end{vmatrix} \\ &= \lambda(\lambda^2 - 4\lambda + 6) - 4(1) \\ &= \lambda^3 - 4\lambda^2 + 6\lambda - 4. \end{aligned}$$

Therefore,

$$\begin{aligned} \det(\lambda I - A) &= \lambda M_{11} - M_{41} \\ &= \lambda(\lambda^3 - 4\lambda^2 + 6\lambda - 4) - (-1) \end{aligned}$$

$$\begin{aligned}
&= \lambda^4 - 4\lambda^3 + 6\lambda^2 - 4\lambda + 1 \\
&= (\lambda - 1)^4.
\end{aligned}$$

The eigenvalue of A is $\lambda = 1$ of multiplicity 4. Note that the rank of the matrix $A - 1 \cdot I$ cannot be 4 as it is a 4×4 matrix with $\det(A - 1 \cdot I) = 0$. So, we can say that the rank of

$$A - I = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ -1 & 4 & -6 & 3 \end{bmatrix}$$

is 3; by noting that the first three rows are linearly independent, or the determinant of the 3×3 principal submatrix of the upper left part is not zero. Hence, the rank of $J - I$ is also 3 and so $J - I$ must have three 1's beyond diagonal entries. It means that

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \square$$

Exercise 4.1.9. Let A be a 5×5 matrix with two distinct eigenvalues: λ of multiplicity 3 and μ of multiplicity 2. Find all possible Jordan canonical forms of A up to permutations of the Jordan blocks.

Exercise 4.1.10. Find the Jordan canonical form for each of the following matrices:

- (1) $\begin{bmatrix} i & 0 \\ 1 & i \end{bmatrix},$
- (2) $\begin{bmatrix} 4 & 1 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 4 \end{bmatrix},$
- (3) $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$

4.2 Generalized eigenvectors

Note that if J is the Jordan canonical form of a matrix A , then they have the same eigenvalues and the same number of linearly independent eigenvectors, but not the same set of them in general. Note that $\mathbf{x}$ is an eigenvector of $J = Q^{-1}AQ$ if and only if $Q\mathbf{x}$ is an eigenvector of A because if $J\mathbf{x} = \lambda\mathbf{x}$, then

$$J = Q^{-1}AQ \Rightarrow QJ\mathbf{x} = AQ\mathbf{x} \Rightarrow \lambda Q\mathbf{x} = A(Q\mathbf{x}).$$

In fact, if J is given, then a basis-change matrix Q is a nonsingular solution of a matrix equation $AQ = QJ$.

The following example illustrates how to determine a basis-change matrix Q when a matrix A and its Jordan canonical form J are given.

Example 4.2.1 (Each Jordan block to each chain of generalized eigenvectors). Let A be a 5×5 matrix similar to a Jordan block of the form

$$Q^{-1}AQ = J = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}.$$

Determine a basis-change matrix $Q = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3 \ \mathbf{x}_4 \ \mathbf{x}_5]$.

Solution: Clearly, two similar matrices A and J have the same eigenvalue λ of multiplicity 5. Since $\text{rank}(J - \lambda I) = 4$, by the rank theorem, $\dim \mathcal{N}(J - \lambda I) = \dim E(\lambda) = 5 - 4 = 1$. In fact, it is easy to see that J has only one linearly independent eigenvector, which is $\mathbf{e}_1 = (1, 0, 0, 0, 0)$. Then as seen above $Q\mathbf{e}_1 = \mathbf{x}_1$ is a linearly independent eigenvector of A . Also, note that the smallest positive integer k such that $(A - \lambda I)^k = (J - \lambda I)^k = \mathbf{0}$ is 5 which is the order of the block J .

To see what the other columns of Q are, we expand the matrix equation $AQ = QJ$ as

$$[A\mathbf{x}_1 \ A\mathbf{x}_2 \ A\mathbf{x}_3 \ A\mathbf{x}_4 \ A\mathbf{x}_5] = [\lambda\mathbf{x}_1 \ \mathbf{x}_1 + \lambda\mathbf{x}_2 \ \mathbf{x}_2 + \lambda\mathbf{x}_3 \ \mathbf{x}_3 + \lambda\mathbf{x}_4 \ \mathbf{x}_4 + \lambda\mathbf{x}_5].$$

By comparing the column vectors, we have

$$\begin{array}{ll} A\mathbf{x}_5 = \mathbf{x}_4 + \lambda\mathbf{x}_5, & \text{or } (A - \lambda I)\mathbf{x}_5 = (A - \lambda I)^1\mathbf{x}_5 = \mathbf{x}_4, \\ A\mathbf{x}_4 = \mathbf{x}_3 + \lambda\mathbf{x}_4, & \text{or } (A - \lambda I)\mathbf{x}_4 = (A - \lambda I)^2\mathbf{x}_5 = \mathbf{x}_3, \\ A\mathbf{x}_3 = \mathbf{x}_2 + \lambda\mathbf{x}_3, & \text{or } (A - \lambda I)\mathbf{x}_3 = (A - \lambda I)^3\mathbf{x}_5 = \mathbf{x}_2, \\ A\mathbf{x}_2 = \mathbf{x}_1 + \lambda\mathbf{x}_2, & \text{or } (A - \lambda I)\mathbf{x}_2 = (A - \lambda I)^4\mathbf{x}_5 = \mathbf{x}_1, \\ A\mathbf{x}_1 = \lambda\mathbf{x}_1, & \text{or } (A - \lambda I)\mathbf{x}_1 = (A - \lambda I)^5\mathbf{x}_5 = \mathbf{0}. \end{array}$$

Note that the vector $\mathbf{x}_5$ satisfies $(A - \lambda I)^5\mathbf{x}_5 = \mathbf{0}$ but $(A - \lambda I)^4\mathbf{x}_5 = \mathbf{x}_1 \neq \mathbf{0}$. However, $(A - \lambda I)^5 = (J - \lambda I)^5 = \mathbf{0}$. Hence, in principle, we need to find only **one vector** (called the rank-5 generalized eigenvector) such that $(A - \lambda I)^4\mathbf{x}_5 \neq \mathbf{0}$, i.e., a solution of $(A - \lambda I)^4\mathbf{x} \neq \mathbf{0}$. Once we get $\mathbf{x}_5$, then all other $\mathbf{x}_i$'s can be obtained as $\mathbf{x}_4 = (A - \lambda I)\mathbf{x}_5$, $\mathbf{x}_3 = (A - \lambda I)\mathbf{x}_4$, $\mathbf{x}_2 = (A - \lambda I)\mathbf{x}_3$, etc.

Such a vector $\mathbf{x}_5$ is called a *generalized eigenvector* of rank 5, and the (ordered) set $\{\mathbf{x}_1, \dots, \mathbf{x}_5\}$ is called a *chain of generalized eigenvectors* belonging to λ . Therefore, the columns of the basis-change matrix Q form a chain of generalized eigenvectors. $\square$

Definition 4.2.2. A nonzero vector $\mathbf{x}$ is said to be a **generalized eigenvector** of A of **rank** k belonging to an eigenvalue λ if

$$(A - \lambda I)^k \mathbf{x} = \mathbf{0} \quad \text{and} \quad (A - \lambda I)^{k-1} \mathbf{x} \neq \mathbf{0}.$$

Note that if $k = 1$, this is the usual definition of an eigenvector. For a generalized eigenvector $\mathbf{x}$ of rank $k \geq 1$ belonging to an eigenvalue λ , define

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{x}, \\
\mathbf{x}_{k-1} &= (A - \lambda I)\mathbf{x}_k &= (A - \lambda I)\mathbf{x}, \\
\mathbf{x}_{k-2} &= (A - \lambda I)\mathbf{x}_{k-1} &= (A - \lambda I)^2\mathbf{x}, \\
&\vdots \\
\mathbf{x}_2 &= (A - \lambda I)\mathbf{x}_3 &= (A - \lambda I)^{k-2}\mathbf{x}, \\
\mathbf{x}_1 &= (A - \lambda I)\mathbf{x}_2 &= (A - \lambda I)^{k-1}\mathbf{x}.
\end{aligned}$$

Definition 4.2.3. The set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ is called a **chain of generalized eigenvectors** belonging to the eigenvalue λ . The eigenvector $\mathbf{x}_1$ is called the **initial eigenvector** of the chain.

The following successive three theorems show that a basis-change matrix Q can be constructed from the chains of generalized eigenvectors initiated from s linearly independent eigenvectors of A , and also justify the invertibility of Q . (A reader may omit reading their proofs below if not interested in details.)

Theorem 4.2.4. A chain of generalized eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ belonging to an eigenvalue λ is linearly independent.

Proof. Let $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_k\mathbf{x}_k = \mathbf{0}$ with constants $c_i, i = 1, \dots, k$. Multiplying both sides (to the left) by $A - \lambda I$, we get

$$\begin{aligned}
c_1(A - \lambda I)\mathbf{x}_1 + c_2(A - \lambda I)\mathbf{x}_2 + \dots + c_k(A - \lambda I)\mathbf{x}_k &= \mathbf{0} \\
\Rightarrow c_2(A - \lambda I)\mathbf{x}_2 + \dots + c_k(A - \lambda I)\mathbf{x}_k &= \mathbf{0} \quad (\because (A - \lambda I)\mathbf{x}_1 = \mathbf{0}).
\end{aligned}$$

Again multiplying above on both sides (to the left) by $A - \lambda I$, we get

$$\begin{aligned}
c_2(A - \lambda I)^2\mathbf{x}_2 + c_3(A - \lambda I)^2\mathbf{x}_3 + \dots + c_k(A - \lambda I)^2\mathbf{x}_k &= \mathbf{0} \\
\Rightarrow c_3(A - \lambda I)^2\mathbf{x}_3 + \dots + c_k(A - \lambda I)^2\mathbf{x}_k &= \mathbf{0} \\
(\because (A - \lambda I)^2\mathbf{x}_2 = (A - \lambda I)((A - \lambda I)\mathbf{x}_1) &= \mathbf{0}).
\end{aligned}$$

Thus, if we multiply (on the left) both sides of this equation by $(A - \lambda I)^{k-1}$, then for $i = 1, \dots, k-1$, we get

$$(A - \lambda I)^{k-1}\mathbf{x}_i = (A - \lambda I)^{k-(i+1)}(A - \lambda I)^i\mathbf{x}_i = \mathbf{0}.$$

Thus, $c_k(A - \lambda I)^{k-1}\mathbf{x}_k = \mathbf{0}$, and, hence, $c_k = 0$. Then our equation has the form

$$c_1\mathbf{x}_1 + \dots + c_{k-1}\mathbf{x}_{k-1} = \mathbf{0}.$$

As done before, multiplying it with $(A - \lambda I)^{k-2}$, we get $c_{k-1} = 0$.

Proceeding successively, one can show that $c_i = 0$ for all $i = 1, \dots, k$. □

Theorem 4.2.5. *The union of chains of generalized eigenvectors of a square matrix A belonging to distinct eigenvalues is linearly independent.*

Proof. For making the proof simple and concise, we prove this theorem for only two distinct eigenvalues. Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ and $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_\ell\}$ be the chains of generalized eigenvectors of A belonging to the eigenvalues λ and μ , respectively, and let $\lambda \neq \mu$. In order to show that the set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{y}_1, \dots, \mathbf{y}_\ell\}$ is linearly independent, let

$$c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k + d_1\mathbf{y}_1 + \dots + d_\ell\mathbf{y}_\ell = \mathbf{0}$$

with constants c_i 's and d_j 's. We multiply both sides of the equation by $(A - \lambda I)^k$ and note that $(A - \lambda I)^k \mathbf{x}_i = \mathbf{0}$ for all $i = 1, \dots, k$. Thus we have

$$(A - \lambda I)^k(d_1\mathbf{y}_1 + d_2\mathbf{y}_2 + \dots + d_\ell\mathbf{y}_\ell) = \mathbf{0}.$$

Again, multiply this equation by $(A - \mu I)^{\ell-1}$ and note that

- $(A - \mu I)^{\ell-1}(A - \lambda I)^k = (A - \lambda I)^k(A - \mu I)^{\ell-1}$
- $(A - \mu I)^{\ell-1}\mathbf{y}_\ell = \mathbf{y}_1$
- $(A - \mu I)^{\ell-1}\mathbf{y}_i = \mathbf{0}$

for $i = 1, \dots, \ell - 1$. Thus we obtain

$$\mathbf{0} = d_\ell(A - \lambda I)^k\mathbf{y}_1.$$

Because $(A - \mu I)\mathbf{y}_1 = \mathbf{0}$ (or $A\mathbf{y}_1 = \mu\mathbf{y}_1$), this reduces to

$$d_\ell(\mu - \lambda)^k\mathbf{y}_1 = \mathbf{0},$$

which implies that $d_\ell = 0$ by the assumption $\lambda \neq \mu$ and $\mathbf{y}_1 \neq \mathbf{0}$. Proceeding successively, one can show that $d_i = 0$, $i = \ell, \ell - 1, \dots, 2, 1$, so we are left with

$$c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k = \mathbf{0}.$$

Since $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ is already linearly independent by Theorem 8.2, $c_i = 0$ for all $i = 1, \dots, k$. Thus the set of generalized eigenvectors $\{\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{y}_1, \dots, \mathbf{y}_\ell\}$ is linearly independent. $\square$

The next step to determine Q such that $AQ = QJ$ is how to choose chains of generalized eigenvectors from a generalized eigenspace, which is defined below, so that the union of the chains is linearly independent.

Definition 4.2.6. Let λ be an eigenvalue of A . The **generalized eigenspace** of A belonging to λ , denoted by K_λ , is the set

$$K_\lambda = \{\mathbf{x} \in \mathbb{C}^n : (A - \lambda I)^p \mathbf{x} = \mathbf{0} \text{ for some positive integer } p\}.$$

It turns out that $\dim K_\lambda$ is the multiplicity of λ , and it contains the usual eigenspace $\mathcal{N}(A - \lambda I)$.

Example 4.2.7 (Basis-change matrix for a triangular matrix). For a matrix

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix},$$

find a basis-change matrix Q so that $Q^{-1}AQ$ is the Jordan canonical matrix.

Solution: *Method 1:* In Example 4.1.7, the Jordan canonical form J of A is determined as

$$J = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

By comparing the column vectors of $AQ = QJ$ with $Q = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3]$, one can get

$$A\mathbf{x}_1 = 2\mathbf{x}_1, \quad A\mathbf{x}_2 = 2\mathbf{x}_2 + \mathbf{x}_1, \quad A\mathbf{x}_3 = 3\mathbf{x}_3.$$

Then (by definition), $\mathbf{x}_1$ and $\mathbf{x}_3$ are eigenvectors of A belonging to $\lambda = 2$ and $\lambda = 3$ respectively.

1. Eigenvector $\mathbf{x}_1$ belonging to $\lambda = 2$ We solve $(A - 2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 0 & 1 & 4 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

From the second and third rows, $-x_3 = 0 \implies x_3 = 0$. Substituting into the first row:

$$x_2 + 4(0) = 0 \implies x_2 = 0.$$

The variable x_1 is free. Choosing $x_1 = 1$, we obtain: $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

2. Eigenvector $\mathbf{x}_3$ belonging to $\lambda = 3$ We solve $(A - 3I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} -1 & 1 & 4 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Here, x_3 is a free variable:

$$\begin{aligned} -x_2 - x_3 &= 0 \implies x_2 = -x_3, \\ -x_1 + x_2 + 4x_3 &= 0 \implies x_1 = x_2 + 4x_3 = -x_3 + 4x_3 = 3x_3. \end{aligned}$$

Choosing $x_3 = 1$, we obtain: $\mathbf{x}_3 = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$.

Method 1: Solving $(A - 2I)\mathbf{x}_2 = \mathbf{x}_1$ Using $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and let $\mathbf{x}_2 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$:

$$\begin{bmatrix} 0 & 1 & 4 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

From each equation:

$$\text{Row 3: } x_3 = 0,$$

$$\text{Row 2: } -x_3 = 0 \implies x_3 = 0,$$

$$\text{Row 1: } 0x_1 + x_2 + 4x_3 = 1 \implies x_2 + 4(0) = 1 \implies x_2 = 1.$$

The entry x_1 does not appear in any equation, so x_1 is a free parameter (let $x_1 = a$): $\mathbf{x}_2 = \begin{bmatrix} a \\ 1 \\ 0 \end{bmatrix}$.

Choosing the simplest scalar $a = 0$ yields: $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Method 2: Solving $(A - 2I)^2\mathbf{x}_2 = \mathbf{0}$ with $(A - 2I)\mathbf{x}_2 \neq \mathbf{0}$ First, compute $(A - 2I)^2$:

$$(A - 2I)^2 = \begin{bmatrix} 0 & 1 & 4 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

The equation $(A - 2I)^2\mathbf{x} = \mathbf{0}$ requires:

$$\begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \implies x_3 = 0.$$

Thus, any vector in the generalized eigenspace has the form $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$.

To ensure $\mathbf{x}$ has rank 2, we require $(A - 2I)\mathbf{x} \neq \mathbf{0}$:

$$(A - 2I) \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} x_2 \\ 0 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \implies x_2 \neq 0.$$

Choosing $x_1 = 0$ and $x_2 = 1$ gives:

$$\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

which automatically recovers the initial eigenvector $\mathbf{x}_1$:

$$\mathbf{x}_1 = (A - 2I)\mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Thus (by any of the method), we have

$$Q = \begin{bmatrix} 1 & a & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{or we can take} \quad Q = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

□

Example 4.2.8 (*Basis-change matrix for a companion matrix*). Find a basis-change matrix Q so that $Q^{-1}AQ = J$ is the Jordan canonical form of the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 4 & -6 & 4 \end{bmatrix}.$$

Solution: *Method 1:* In Example 4.1.8, we computed

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now, one can find a basis-change matrix $Q = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3 \ \mathbf{x}_4]$ by comparing the column vectors of $AQ = QJ$:

$$A\mathbf{x}_1 = \mathbf{x}_1, \quad A\mathbf{x}_2 = \mathbf{x}_2 + \mathbf{x}_1, \quad A\mathbf{x}_3 = \mathbf{x}_3 + \mathbf{x}_2, \quad A\mathbf{x}_4 = \mathbf{x}_4 + \mathbf{x}_3.$$

By computing an eigenvector of A belonging to $\lambda = 1$, one can get $\mathbf{x}_1 = (1, 1, 1, 1)$. The equation $A\mathbf{x}_2 = \mathbf{x}_2 + \mathbf{x}_1$ gives $\mathbf{x}_2 = (a, a + 1, a + 2, a + 3)$ for any a . Take $\mathbf{x}_2 = (0, 1, 2, 3)$. Similarly, from equations $A\mathbf{x}_3 = \mathbf{x}_3 + \mathbf{x}_2$ and $A\mathbf{x}_4 = \mathbf{x}_4 + \mathbf{x}_3$, one can get $\mathbf{x}_3 = (b, b, b + 1, b + 3)$ for any b and set $\mathbf{x}_3 = (0, 0, 1, 3)$, and successively one can take $\mathbf{x}_4 = (0, 0, 0, 1)$. We conclude that

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{bmatrix}.$$

One may check $AQ = QJ$ by a direct matrix multiplication.

Exercise 4.2.9 (From above Examples). Find a full set of generalized eigenvectors of the following matrices:

- (1) $\begin{bmatrix} i & 0 \\ 1 & i \end{bmatrix},$
- (2) $\begin{bmatrix} 4 & 1 & 2 \\ 0 & 4 & 2 \\ 0 & 0 & 4 \end{bmatrix},$

4.3 Cayley-Hamilton theorem

Let $f(x) = a_mx^m + a_{m-1}x^{m-1} + \cdots + a_1x + a_0$ be a polynomial, and let A be an $n \times n$ square matrix. The matrix defined by

$$f(A) = a_mA^m + a_{m-1}A^{m-1} + \cdots + a_1A + a_0I_n$$

is called a **matrix polynomial** of A .

For example, if $f(x) = x^2 - 2x + 2$ and $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, then

$$\begin{aligned} f(A) &= A^2 - 2A + 2I_2 \\ &= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} - 2 \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + 2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}. \end{aligned}$$

Exercise 4.3.1. Let λ be an eigenvalue of a matrix A . For any polynomial $f(x)$, show that $f(\lambda)$ is an eigenvalue of the matrix polynomial $f(A)$.

The following theorem says that **every matrix satisfies its characteristic equation**.

Theorem 4.3.2 (Cayley-Hamilton). For any $n \times n$ matrix A , if $f(\lambda) = \det(\lambda I - A)$ is the characteristic polynomial of A , then $f(A) = \mathbf{0}$.

Proof. First assume that A is the diagonal matrix given by

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

The characteristic polynomial $f(\lambda)$ of A is

$$f(\lambda) = \det(\lambda I - A) = \det \begin{bmatrix} \lambda - \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda - \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda - \lambda_n \end{bmatrix} = \prod_{j=1}^n (\lambda - \lambda_j).$$

For any positive integer k , powers of the diagonal matrix A are obtained by raising the diagonal entries to that power:

$$A^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix}.$$

Hence, for any polynomial $f(x) = c_mx^m + c_{m-1}x^{m-1} + \cdots + c_1x + c_0$, the matrix polynomial $f(A)$ is

$$f(A) = c_mA^m + c_{m-1}A^{m-1} + \cdots + c_1A + c_0I = \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix}.$$

Since each λ_i for $i = 1, \dots, n$ is a root of the scalar polynomial $f(\lambda)$, we have $f(\lambda_i) = 0$. Substituting these values back into $f(A)$ yields

$$f(A) = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} = \mathbf{0}.$$

For an arbitrary square matrix A , let its Jordan canonical form be

$$J = \begin{bmatrix} J_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_s \end{bmatrix} = Q^{-1}AQ.$$

Since $A^k = QJ^kQ^{-1}$, we have $f(A) = Qf(J)Q^{-1}$. Thus, $f(A) = 0 \Leftrightarrow f(J) = 0$. Since

$$J^k = \begin{bmatrix} J_1^k & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_s^k \end{bmatrix} \quad \text{and} \quad f(J) = \begin{bmatrix} f(J_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & f(J_s) \end{bmatrix},$$

to show that $f(J) = 0$, it is sufficient to show that $f(J_i) = \mathbf{0}$ for each Jordan block J_i . Let $J_i = \lambda_0 I + N$ with eigenvalue λ_0 of multiplicity m , in which $N^m = \mathbf{0}$. Since the determinant of similar matrices is same, $f(\lambda) = \det(\lambda I - A) = \det(\lambda I - J)$. Let $\det(\lambda I - J) = (\lambda - \lambda_0)^m g(\lambda)$ for some $g(\lambda)$. Then, we have

$$f(J_i) = (J_i - \lambda_0 I)^m g(J_i) = (\lambda_0 I + N - \lambda_0 I)^m g(J_i) = N^m g(J_i) = \mathbf{0} \quad g(J_i) = \mathbf{0}.$$

□

Example 4.3.3. The characteristic polynomial of

$$A = \begin{bmatrix} 3 & 6 & 6 \\ 0 & 2 & 0 \\ -3 & -12 & -6 \end{bmatrix}$$

is $f(\lambda) = \det(\lambda I - A) = \lambda^3 + \lambda^2 - 6\lambda$, and

$$\begin{aligned} f(A) &= A^3 + A^2 - 6A \\ &= \begin{bmatrix} 27 & 78 & 54 \\ 0 & 8 & 0 \\ -27 & -102 & -54 \end{bmatrix} + \begin{bmatrix} -9 & -42 & -18 \\ 0 & 4 & 0 \\ 9 & 30 & 18 \end{bmatrix} \\ &\quad - 6 \begin{bmatrix} 3 & 6 & 6 \\ 0 & 2 & 0 \\ -3 & -12 & -6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

□

The Cayley-Hamilton theorem can be used to find the inverse of a nonsingular matrix. If $f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0$ is the characteristic polynomial of a matrix A , then

$$\mathbf{0} = f(A) = A^n + a_{n-1}A^{n-1} + \dots + a_1A + a_0I,$$

$$\text{or } -a_0I = (A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_1I)A.$$

Since $a_0 = f(0) = \det(0I - A) = \det(-A) = (-1)^n \det A$, A is nonsingular if and only if $a_0 = (-1)^n \det A \neq 0$. Therefore, if A is nonsingular,

$$A^{-1} = -\frac{1}{a_0}(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_1I).$$

Example 4.3.4 (Compute A^{-1} by the Cayley-Hamilton theorem). The characteristic polynomial of the matrix

$$A = \begin{bmatrix} 4 & 2 & -2 \\ -5 & 3 & 2 \\ -2 & 4 & 1 \end{bmatrix}$$

is $f(\lambda) = \det(\lambda I_3 - A) = \lambda^3 - 8\lambda^2 + 17\lambda - 10$, and the Cayley-Hamilton theorem yields

$$A^3 - 8A^2 + 17A = 10I_3.$$

Hence

$$\begin{aligned} A^{-1} &= \frac{1}{10}(A^3 - 8A^2 + 17I_3) \\ &= \frac{1}{10} \begin{bmatrix} 10 & 6 & -6 \\ -39 & 7 & 18 \\ -30 & 12 & 13 \end{bmatrix} - \frac{8}{10} \begin{bmatrix} 4 & 2 & -2 \\ -5 & 3 & 2 \\ -2 & 4 & 1 \end{bmatrix} + \frac{17}{10} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} -5 & -10 & 10 \\ 1 & 0 & 2 \\ -14 & -20 & 22 \end{bmatrix}. \end{aligned} \quad \square$$

Exercise 4.3.5. Let A and B be square matrices, not necessarily of the same order, and let $f(\lambda) = \det(\lambda I - A)$ be the characteristic polynomial of A . Show that $f(B)$ is invertible if and only if A has no eigenvalue in common with B .

The Cayley-Hamilton theorem can also be used to simplify the calculation of matrix polynomials. Let $p(\lambda)$ be any polynomial and let $f(\lambda)$ be the characteristic polynomial of a square matrix A . A theorem of algebra tells us that there are polynomials $q(\lambda)$ and $r(\lambda)$ such that

$$p(\lambda) = q(\lambda)f(\lambda) + r(\lambda),$$

where the degree of $r(\lambda)$ is less than the degree of $f(\lambda)$. Then

$$p(A) = q(A)f(A) + r(A).$$

By the Cayley-Hamilton theorem, $f(A) = \mathbf{0}$ and

$$p(A) = r(A).$$

Thus, the problem of evaluating a polynomial of an $n \times n$ matrix A or in particular a power A^k can be reduced to the problem of evaluating a matrix polynomial of degree less than n .

Example 4.3.6. The characteristic polynomial of the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is $f(\lambda) = \lambda^2 - 2\lambda - 3$. Let $p(\lambda) = \lambda^4 - 7\lambda^3 - 3\lambda^2 + \lambda + 4$ be a polynomial. A division by $f(\lambda)$ gives that

$$p(\lambda) = (\lambda^2 - 5\lambda - 10)f(\lambda) - 34\lambda - 26.$$

Therefore

$$\begin{aligned} p(A) &= (A^2 - 5A - 10I)f(A) - 34A - 26I \\ &= -34A - 26I \\ &= -34 \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} - 26 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -60 & -68 \\ -68 & -60 \end{bmatrix}. \end{aligned} \quad \square$$

4.4 The minimal polynomial of a matrix

Let A be a square matrix of order n and let

$$f(\lambda) = \det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = \prod_{i=1}^s (\lambda - \lambda_i)^{m_i}$$

be the characteristic polynomial of A , where m_{λ_i} is the multiplicity of the eigenvalue λ_i . Then, the Cayley–Hamilton theorem says that

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1A + a_0I = \prod_{i=1}^s (A - \lambda_i I)^{m_i} = \mathbf{0}.$$

The **minimal polynomial** of a matrix A is the monic polynomial $m(\lambda) = \sum_{i=0}^m c_i \lambda^i$ of smallest degree m such that

$$m(A) = \sum_{i=0}^m c_i A^i = \mathbf{0}.$$

Proposition 4.4.1. *The minimal polynomial of a matrix A divides any polynomial $p(\lambda)$ such satisfying $p(A) = \mathbf{0}$.*

In particular, the minimal polynomial of A divides its characteristic polynomial.

Proof. Let $m(\lambda)$ be the minimal polynomial of the matrix A and $p(\lambda)$ be a polynomial of A satisfying $p(A) = \mathbf{0}$.

Let $p(\lambda) = q(\lambda)m(\lambda) + r(\lambda)$, where the degree of $r(\lambda)$ is less than the degree of $m(\lambda)$. Since we are given that $p(A) = \mathbf{0}$ and by the definition of minimal polynomial $m(A) = \mathbf{0}$, we have

$$\mathbf{0} = p(A) = q(A)m(A) + r(A) = r(A)$$

Since the degree of $r(\lambda)$ is less than the degree of $m(\lambda)$, the minimality of $m(\lambda)$ implies that $r(\lambda) = 0$. Hence,

$$p(\lambda) = q(\lambda)m(\lambda) \text{ or } m(\lambda) \mid p(\lambda).$$

Thus, minimal polynomial of A divides any polynomial $p(\lambda)$ such satisfying $p(A) = \mathbf{0}$.

In particular, by the Cayley–Hamilton theorem, the minimal polynomial $m(\lambda)$ divides the characteristic polynomial. $\square$

Since the minimal polynomial $m(\lambda)$ divides the characteristic polynomial $f(\lambda) = \det(\lambda I - A) = \prod_{i=1}^s (\lambda - \lambda_i)^{m_i}$, we have

$$m(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{k_i}$$

with $k_i \leq m_i$. For example, the characteristic polynomial of the $n \times n$ zero matrix is λ^n and its minimal polynomial is just λ .

Clearly, any two similar matrices have the same minimal polynomial because

$$\prod_{i=1}^s (Q^{-1}AQ - \lambda_i I)^{k_i} = Q^{-1} \left(\prod_{i=1}^s (A - \lambda_i I)^{k_i} \right) Q$$

for any invertible matrix Q .

Example 4.4.2 (For a diagonal matrix).

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix},$$

its characteristic polynomial is $f(\lambda) = (\lambda - 2)^3(\lambda - 5)^2$. However, for the polynomial $m(\lambda) = (\lambda - 2)(\lambda - 5)$, the matrix $m(A)$ is

$$m(A) = (A - 2I)(A - 5I) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -3 & 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{0}.$$

Hence, the minimal polynomial of A is $m(\lambda) = (\lambda - 2)(\lambda - 5)$. $\square$

Example 4.4.3 (The minimal polynomial of a diagonal matrix). **(1)** Any diagonal matrix of the form $\lambda_0 I$ has the minimal polynomial $\lambda - \lambda_0$. In particular, the minimal polynomial of the zero matrix is the monomial λ , and the minimal polynomial of the identity matrix is the monomial $\lambda - 1$.

(2) If an $n \times n$ matrix A has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$, then its minimal polynomial coincides with the characteristic polynomial $f(\lambda) = \prod_{i=1}^n (\lambda - \lambda_i)$.

Proof. Let $m(\lambda)$ be the minimal polynomial of the diagonal matrix D . We know that $m(\lambda) \mid f(\lambda)$. We claim that $m(\lambda) = f(\lambda)$. Since $f(\lambda)$ is a product of distinct linear factors $(\lambda - \lambda_i)$, if $m(\lambda) \neq f(\lambda)$, then $m(\lambda)$ must omit at least one factor $(\lambda - \lambda_j)$. We know that $m(D) = \mathbf{0}$. But the (j, j) -entry of $m(D) = \prod_{i \neq j} (D - \lambda_i I)$ is equal to $\prod_{i \neq j} (\lambda_j - \lambda_i) \neq 0$. Hence, the minimal polynomial is the same as the characteristic polynomial. $\square$

Example 4.4.4 (The minimal polynomial of a Jordan matrix J with one or two blocks). **(1)**

For any 5×5 matrix A similar to a Jordan block of the form

$$Q^{-1}AQ = J = \begin{bmatrix} \lambda_0 & 1 & 0 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_0 \end{bmatrix},$$

its minimal polynomial is equal to the characteristic polynomial $f(\lambda) = (\lambda - \lambda_0)^5$, because $(J - \lambda_0 I)^4 \neq \mathbf{0}$ but $(J - \lambda_0 I)^5 = \mathbf{0}$.

- (2) Consider a matrix J having two Jordan blocks belonging to a single eigenvalue λ , say

$$J = \begin{bmatrix} J_1 & \mathbf{0} \\ \mathbf{0} & J_2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \lambda_0 & 1 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & \lambda_0 \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} \lambda_0 & 1 & 0 \\ 0 & \lambda_0 & 1 \\ 0 & 0 & \lambda_0 \end{bmatrix} \end{bmatrix}.$$

For J_1 of size 4, we have

$$(J_1 - \lambda_0 I_4)^3 \neq \mathbf{0}, \quad (J_1 - \lambda_0 I_4)^4 = \mathbf{0}.$$

So the minimal polynomial of J_1 alone is: $m_{J_1}(\lambda) = (\lambda - \lambda_0)^4$.

For J_2 of size 3, we have

$$(J_2 - \lambda_0 I_3)^2 \neq \mathbf{0}, \quad (J_2 - \lambda_0 I_3)^3 = \mathbf{0}.$$

So the minimal polynomial of J_2 alone is: $m_{J_2}(\lambda) = (\lambda - \lambda_0)^3$.

Hence, the minimal polynomial of J is

$$m(\lambda) = \text{lcm}((\lambda - \lambda_0)^4, (\lambda - \lambda_0)^3) = (\lambda - \lambda_0)^{\max(4,3)} = (\lambda - \lambda_0)^4.$$

- (3) For a matrix J having two Jordan blocks belonging to two **different** eigenvalues $\lambda_0 \neq \lambda_1$ respectively, say

$$J = \begin{bmatrix} J_1 & \mathbf{0} \\ \mathbf{0} & J_2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \lambda_0 & 1 & 0 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_0 \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix} \end{bmatrix},$$

the minimal polynomial of J is a product of the minimal polynomials of J_1 and J_2 , which is $(\lambda - \lambda_0)^5(\lambda - \lambda_1)^3$.

By a method similar to the above example, one can have the following theorem.

Theorem 4.4.5. For any $n \times n$ matrix A or its Jordan canonical matrix J , its minimal polynomial is $\prod_{i=1}^t (\lambda - \lambda_i)^{k_i}$, where $\lambda_1, \dots, \lambda_t$ are the distinct eigenvalues of A and k_i is the order of the largest Jordan block in J belonging to the eigenvalue λ_i .

Since for a diagonal matrix, every block is of size 1, we have the following corollary for the minimal polynomial of a diagonalizable matrix.

Corollary 4.4.6. A matrix A is diagonalizable if and only if its minimal polynomial is equal to $\prod_{i=1}^t (\lambda - \lambda_i)$, where $\lambda_1, \dots, \lambda_t$ are the distinct eigenvalues of A .

Example 4.4.7 (Computing the minimal polynomial). Compute the minimal polynomial of A for

$$(1) A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}, \quad (2) A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 4 & -6 & 4 \end{bmatrix}.$$

Solution: (1) Since A is triangular, its eigenvalues are $\lambda_1 = \lambda_2 = 2, \lambda_3 = 3$. But, $\text{rank}(A - 2I) = 2$ and so A is not diagonalizable. Hence, its minimal polynomial is $m(A) = (\lambda - 2)^2(\lambda - 3)$.

(2) Recalling Example 4.1.8, we know that the eigenvalue of A is $\lambda = 1$ of multiplicity 4, and $\text{rank}(A - I) = 3$. It implies that the Jordan canonical form of A is

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and the minimal polynomial of A is $m(A) = (\lambda - 1)^4$, by the above theorem. $\square$

Example 4.4.8. Compute the minimal polynomial of A for

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ -1 & 1 & 0 & 3 \end{bmatrix}.$$

Solution: First **show that** the characteristic polynomial of A is $f(\lambda) = (\lambda - 2)^4$. Note that $\text{rank}(A - 2I) = 2$. So, this rank 2 is coming from a block of order 3 or two blocks of order 2 each in the 4×4 matrix. However, $\text{rank}(A - 2I)^2 = 1$ and this implies that there must be one block of order 3; for if there were two blocks of order 2, then the rank of $(A - 2I)^2$ would be 0. Thus, the Jordan canonical form J of A is

$$J = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

So, the minimal polynomial of A is $m(A) = (\lambda - 2)^3$, by the above theorem. $\square$

Exercise 4.4.9. In Example 4.1.5, we have seen that there are seven possible (nonsimilar) Jordan canonical matrices of order 5 that have a single eigenvalue. Compute the minimal polynomial of each of them.

4.5 Quadratic Forms

4.5.1 Basic properties of quadratic forms

A **quadratic equation** in two variables x and y is an equation of the form

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0,$$

in which the left-hand side consists of a constant term f , a linear form $dx + ey$, and a quadratic form $ax^2 + 2bxy + cy^2$. Note that this quadratic form may be written in matrix notation as

$$ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^T A \mathbf{x},$$

where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Note also that the matrix A is taken to be a (real) symmetric matrix.

Geometrically, the solution set of a quadratic equation in x and y usually represents a *conic section*, such as an ellipse, a parabola or a hyperbola in the xy -plane.

Definition 4.5.1. (1) A **linear form** on the Euclidean space $\mathbb{R}^n$ is a polynomial of degree 1 in n variables $x_1, x_2, \dots, x_n$ of the form

$$\mathbf{b}^T \mathbf{x} = \sum_{i=1}^n b_i x_i,$$

where $\mathbf{x} = [x_1 \dots x_n]^T$ and $\mathbf{b} = [b_1 \dots b_n]^T$ in $\mathbb{R}^n$.

(2) A **quadratic equation** on $\mathbb{R}^n$ is an equation in n variables $x_1, x_2, \dots, x_n$ of the form

$$f(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n b_i x_i + c = 0,$$

where a_{ij} , b_j and c are real constants. In matrix form, it can be written as

$$f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0,$$

where $A = [a_{ij}]$, $\mathbf{x} = [x_1 \dots x_n]^T$ and $\mathbf{b} = [b_1 \dots b_n]^T$ in $\mathbb{R}^n$.

(3) A **quadratic form** on $\mathbb{R}^n$ is a (homogeneous) polynomial of degree 2 in n variables $x_1, x_2, \dots, x_n$ of the form

$$q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j,$$

where $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T \in \mathbb{R}^n$ and $A = [a_{ij}]$ is a real $n \times n$ symmetric matrix.

It is also possible to define a linear form and a quadratic form on the Euclidean complex n -space $\mathbb{C}^n$ instead of the Euclidean real n -space $\mathbb{R}^n$.

The real quadratic form on $\mathbb{R}^n$ and the complex quadratic form on $\mathbb{C}^n$ can be denoted simultaneously as $q(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x} \rangle$ by using the dot product on the real n -space $\mathbb{R}^n$ or on the complex n -space $\mathbb{C}^n$.

Remark 4.5.2. A quadratic equation $f(\mathbf{x})$ is said to be **consistent** if it has a solution, i.e., there is a vector $\mathbf{x} \in \mathbb{R}^n$ such that $f(\mathbf{x}) = 0$. Otherwise, it is said to be **inconsistent**. For instance, the equation $2x^2 + 3y^2 = -1$ in $\mathbb{R}^2$ is inconsistent. In what follows, we will consider only consistent equations.

The solution set of a consistent quadratic equation $f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0$ is a level surface in $\mathbb{R}^n$, that is, a curved surface that can be parameterized in $n - 1$ variables. In particular, if $n = 2$, the solution set of a quadratic equation is called a **quadratic curve**, or more commonly a **conic section**. When $n = 3$, it is called a **quadratic surface**, which is an ellipsoid, a paraboloid or a hyperboloid.

Example 4.5.3 (The standard three types of conic sections). (1) (circle or ellipse) $\frac{x^2}{a^2} + \frac{y^2}{b^2} =$

$$1 \text{ with } A = \begin{bmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{bmatrix}.$$

(2) (hyperbola) $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ or $\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$ with

$$A = \begin{bmatrix} \frac{1}{a^2} & 0 \\ 0 & -\frac{1}{b^2} \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} -\frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{bmatrix}.$$

(3) (parabola) $x^2 = ay$ or $y^2 = bx$ with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Example 4.5.4 (The standard four types of quadratic surfaces). (1) (ellipsoids) $\frac{x^2}{a^2} + \frac{y^2}{b^2} +$

$$\frac{z^2}{c^2} = 1 \text{ with } A = \begin{bmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & \frac{1}{c^2} \end{bmatrix}.$$

(2) (hyperboloids of one or two sheets) $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$ (of one sheet) or $-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ (of two sheets) with

$$A = \begin{bmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & -\frac{1}{c^2} \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} -\frac{1}{a^2} & 0 & 0 \\ 0 & -\frac{1}{b^2} & 0 \\ 0 & 0 & \frac{1}{c^2} \end{bmatrix}.$$

(3) (cones) $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$ with $A = \begin{bmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & -\frac{1}{c^2} \end{bmatrix}.$

(4) (paraboloids; elliptic or hyperbolic) $\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z}{c}$ (elliptic) or $\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c}$, $c > 0$ (hyperbolic) with

$$A = \begin{bmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & -\frac{1}{b^2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

• **Ellipsoid:**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

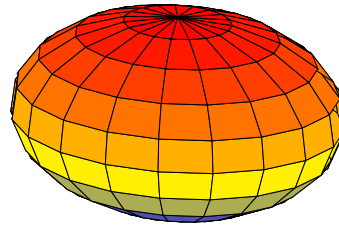


Figure 4.1: Ellipsoid

• **Hyperboloid with 1 sheet:**

$$x^2 + y^2 - z^2 = 1.$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

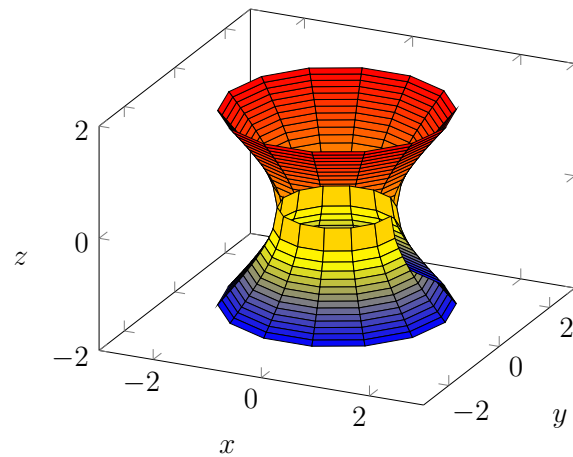


Figure 4.2: Hyperboloid with 1 sheet

• **Hyperboloid with 2 sheets:**

$$x^2 - y^2 - z^2 = 1.$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

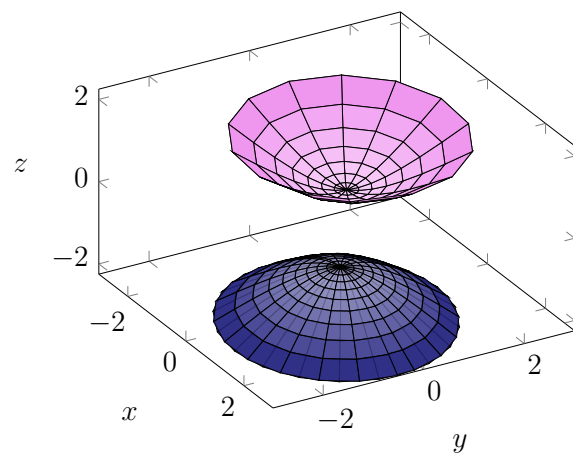


Figure 4.3: Hyperboloid with 2 sheets

• **Cone:**

$$x^2 + y^2 = z^2.$$

$$ax^2 + by^2 = z^2 \quad (a \text{ and } b \text{ are positive}).$$

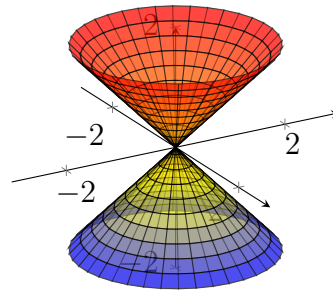


Figure 4.4: Cone

Exercise 4.5.5. Find the symmetric matrices representing the quadratic forms

- (1) $9x_1^2 - x_2^2 + 4x_3^2 + 6x_1x_2 - 8x_1x_3 + 2x_2x_3$,
- (2) $xy + xz + yz$,
- (3) $x_1^2 + x_2^2 - x_3^2 - x_4^2 + 2x_1x_2 - 10x_1x_4 + 4x_3x_4$,
- (4) $-y^2 - 2z^2 + 4xy + 8xz - 14yz$.

Solution. The matrices associated with the above quadratic forms are:

$$(1) \begin{pmatrix} 9 & 3 & -4 \\ 3 & -1 & 1 \\ -4 & 1 & 4 \end{pmatrix}.$$

$$(2) \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

$$(3) \begin{pmatrix} 1 & 1 & 0 & -5 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 2 \\ -5 & 0 & 2 & -1 \end{pmatrix}.$$

$$(4) \begin{pmatrix} 0 & 2 & 4 \\ 2 & -1 & -7 \\ 4 & -7 & -2 \end{pmatrix}.$$

□

• **Paraboloid:**

$$z = x^2 + y^2.$$

$$z = ax^2 + by^2 \quad (a, b > 0).$$

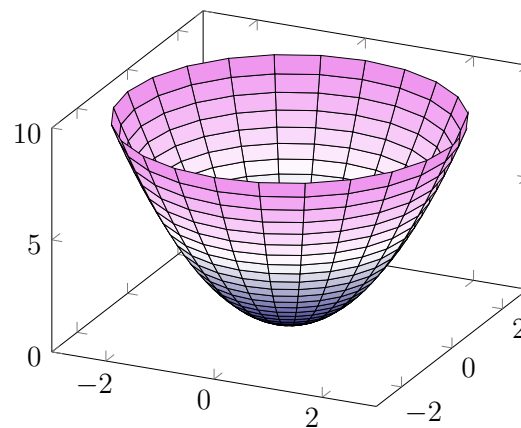


Figure 4.5: Elliptic Paraboloid

• **Hyperbolic Paraboloid:**

$$z = x^2 - y^2.$$

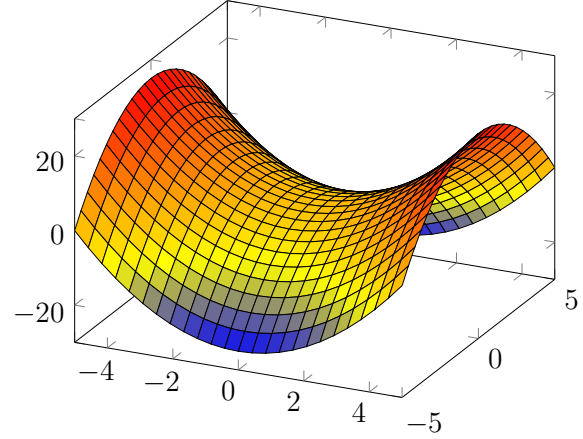


Figure 4.6: Hyperbolic Paraboloid

4.6 Diagonalization of quadratic forms

A quadratic form on $\mathbb{R}^n$ without a linear form may be written as the sum of two parts:

$$q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{i < j} a_{ij} x_i x_j,$$

in which the first part $\sum_{i=1}^n a_{ii} x_i^2$ is called the **(perfect) square terms** and the second part $\sum_{i \neq j} a_{ij} x_i x_j$ is called the **cross-product terms**. The quadratic form $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ can be transformed into a new quadratic form without the cross-product terms by a suitable change of variables. It can be done by computing the eigenvalues of A and their associated eigenvectors. In fact, the symmetric matrix A can be orthogonally diagonalized, i.e., there exists an orthogonal matrix P such that

$$P^T A P = P^{-1} A P = D = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \lambda_2 & \\ & & \ddots \\ \mathbf{0} & & & \lambda_n \end{bmatrix}.$$

Here, the diagonal entries λ_i 's are the eigenvalues of A and the column vectors of P are their associated eigenvectors of A . Now, by setting $\mathbf{x} = P\mathbf{y}$, (that is, by a change of variables), we have

$$\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T (P^T A P) \mathbf{y} = \mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

which is a quadratic form without the cross-product terms.

Theorem 4.6.1 (The principal axes theorem). *Let $\mathbf{x}^T A \mathbf{x}$ be a quadratic form in $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T \in \mathbb{R}^n$ for a symmetric matrix A . Then, there is a change of coordinates of $\mathbf{x}$ into $\mathbf{y} = P^T \mathbf{x} = [y_1 \ y_2 \ \dots \ y_n]^T$ such that*

$$\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

where P is an orthogonal matrix and $P^T AP = P^{-1}AP = D$ is diagonal.

Clearly, the columns of the matrix P in the above theorem form an orthonormal basis for $\mathbb{R}^n$ and it is called the **principal axes** of the quadratic form. The vector $\mathbf{y}$ is just the coordinate expression of $\mathbf{x}$ with respect to the principal axes.

Example 4.6.2 (Via a diagonalization of a quadratic form). Determine the conic section $3x^2 + 2xy + 3y^2 - 8 = 0$ on $\mathbb{R}^2$.

Solution: In matrix form, it is

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 8.$$

The matrix $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ has eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 4$ with associated unit eigenvectors

$$\mathbf{v}_1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right) \quad \text{and} \quad \mathbf{v}_2 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right),$$

respectively, which form an orthonormal basis β . If α denotes the standard basis, then the basis-change matrix

$$P = [\text{id}]_{\beta}^{\alpha} = [\mathbf{v}_1 \ \mathbf{v}_2] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \cos 45^\circ & \sin 45^\circ \\ -\sin 45^\circ & \cos 45^\circ \end{bmatrix},$$

which is a rotation through 45° in the clockwise direction such that $P^T = P^{-1}$. It gives a change of coordinates, $\mathbf{x} = P\mathbf{y}$, i.e.,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}y' \\ -\frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}y' \end{bmatrix}.$$

It implies that

$$\begin{aligned} 3x^2 + 2xy + 3y^2 &= \mathbf{x}^T A \mathbf{x} = \mathbf{y}^T P^T A P \mathbf{y} \\ &= \mathbf{y}^T \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \mathbf{y} = 2(x')^2 + 4(y')^2 = 8, \end{aligned}$$

or

$$\frac{(x')^2}{4} + \frac{(y')^2}{2} = 1,$$

which is an ellipse with the principal axes $\mathbf{v}_1 = P^T \mathbf{e}_1$ and $\mathbf{v}_2 = P^T \mathbf{e}_2$. □

To determine the type of a quadratic form, we have the following definition for a symmetric matrix A or a quadratic form $\mathbf{x}^T A \mathbf{x}$.

Definition 4.6.3. Let $A = [a_{ij}] \in M_{n \times n}(\mathbb{R})$ be a symmetric matrix and let $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. Then, the matrix A , or a quadratic form $\mathbf{x}^T A \mathbf{x}$, is said to be

- (1) **positive definite** if $\mathbf{x}^T A \mathbf{x} = \sum_{i,j} a_{ij} x_i x_j > 0$ for all nonzero $\mathbf{x}$,
- (2) **positive semidefinite** if $\mathbf{x}^T A \mathbf{x} = \sum_{i,j} a_{ij} x_i x_j \geq 0$ for all $\mathbf{x}$,
- (3) **negative definite** if $\mathbf{x}^T A \mathbf{x} = \sum_{i,j} a_{ij} x_i x_j < 0$ for all nonzero $\mathbf{x}$,

- (4) **negative semidefinite** if $\mathbf{x}^T A \mathbf{x} = \sum_{i,j} a_{ij} x_i x_j \leq 0$ for all $\mathbf{x}$,
 (5) **indefinite** if $\mathbf{x}^T A \mathbf{x}$ takes both positive and negative values.

For example, the real symmetric matrix

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

is positive definite, because the quadratic form satisfies

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 \end{bmatrix} \\ &= x_1(2x_1 - x_2) + x_2(-x_1 + 2x_2 - x_3) + x_3(-x_2 + 2x_3) \\ &= 2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2 \\ &= x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 > 0 \end{aligned}$$

unless $x_1 = x_2 = x_3 = 0$.

The following characterizations follow from the principal axes theorem.

Corollary 4.6.4. *A real symmetric or a Hermitian matrix A is*

- (1) **positive definite** if and only if all the eigenvalues of A are positive,
- (2) **positive semidefinite** if and only if all the eigenvalues of A are nonnegative,
- (3) **negative definite** if and only if all the eigenvalues of A are negative,
- (4) **negative semidefinite** if and only if all the eigenvalues of A are nonpositive,
- (5) **indefinite** if and only if A takes both positive and negative eigenvalues.

Note that if A is positive definite, $\det A > 0$ (as the product of all eigenvalues). If the eigenvalues of A are all negative, then $-A$ must be positive definite and consequently A must be negative definite. If A has eigenvalues that differ in sign, then A is **indefinite**. Indeed, if λ_1 is a positive eigenvalue of A and $\mathbf{x}_1$ is an eigenvector belonging to λ_1 , then

$$q(\mathbf{x}_1) = \langle \mathbf{x}_1, A\mathbf{x}_1 \rangle = \lambda_1 \langle \mathbf{x}_1, \mathbf{x}_1 \rangle = \lambda_1 \|\mathbf{x}_1\|^2 > 0,$$

and if λ_2 is a negative eigenvalue with eigenvector $\mathbf{x}_2$, then

$$q(\mathbf{x}_2) = \langle \mathbf{x}_2, A\mathbf{x}_2 \rangle = \lambda_2 \langle \mathbf{x}_2, \mathbf{x}_2 \rangle = \lambda_2 \|\mathbf{x}_2\|^2 < 0.$$

If A is definite, then $\mathbf{0}$ is the only critical point of a quadratic form $q(\mathbf{x}) = \langle \mathbf{x}, A\mathbf{x} \rangle$, and $q(\mathbf{0}) = 0$ is the *global minimum* if A is positive definite and the *global maximum* if A is negative definite. If A is indefinite, then $\mathbf{0}$ is a *saddle point*.

4.7 A classification of level surfaces

The geometric type of the level surface of a quadratic equation $\mathbf{x}^T A \mathbf{x} = 0$ depends on the signs of the eigenvalues of A . In fact, it is completely determined by the numbers of positive,

negative and zero eigenvalues of A .

Definition 4.7.1. The **inertia** of a Hermitian (or a symmetric) matrix A is a triple of integers denoted by $\text{In}(A) = (p, q, k)$, where p , q and k are the numbers of positive, negative and zero eigenvalues of A , respectively.

The inertia $\text{In}(A)$ determines the geometric type of the quadratic surface $\mathbf{x}^T A \mathbf{x} = 0$ on $\mathbb{R}^n$ in the following sense. Since $\text{In}(-A) = (q, p, k)$ if $\text{In}(A) = (p, q, k)$ and the equation $\mathbf{x}^T A \mathbf{x} = c$ is inconsistent if $p = 0$ and $c > 0$, it suffices to consider the cases of $c \geq 0$ and $p > 0$. Excluding those inconsistent cases, we have the following characterization of the solution sets for $n = 2$ and 3:

For $n = 2$, there are only three possible cases for $\text{In}(A)$:

$\text{In}(A)$ (p, q, k)	The solution of $\mathbf{x}^T A \mathbf{x} = c$	
	$c > 0$	$c = 0$
$(2, 0, 0)$	ellipse	a point
$(1, 1, 0)$	hyperbola	two lines crossing at $\mathbf{0}$
$(1, 0, 1)$	two parallel lines	a line

For $n = 3$, there are six possibilities:

$\text{In}(A)$ (p, q, k)	The solution of $\mathbf{x}^T A \mathbf{x} = c$	
	$c > 0$	$c = 0$
$(3, 0, 0)$	ellipsoid	a point
$(2, 1, 0)$	one-sheeted hyperboloid	elliptic cone
$(2, 0, 1)$	elliptic cylinder	a line
$(1, 2, 0)$	two-sheeted hyperboloid	elliptic cone
$(1, 1, 1)$	hyperbolic cylinder	two planes crossing in a line
$(1, 0, 2)$	two parallel planes	a plane

In general, for an $n \times n$ symmetric matrix A , $\text{In}(A)$ will have $n(n+1)/2$ possibilities, each characterizing a different geometric type of a quadratic form. For example, if $\text{In}(A) = (n, 0, 0)$ and $c > 0$, i.e., the eigenvalues of A are all positive, then the quadratic form describes an ellipsoid in $\mathbb{R}^n$, etc.

Example 4.7.2 (*The inertia determines the geometric type of the quadratic form*). Determine the quadratic surface for $2xy + 2xz = 1$ on $\mathbb{R}^3$.

Solution: The matrix for the given quadratic form is

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

and the eigenvalues of A can be found to be $\lambda_1 = \sqrt{2}$, $\lambda_2 = -\sqrt{2}$, $\lambda_3 = 0$, with associated orthonormal eigenvectors

$$\mathbf{v}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{2}, \frac{1}{2} \right), \quad \mathbf{v}_2 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{2}, \frac{1}{2} \right), \quad \mathbf{v}_3 = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right),$$

respectively. Hence, an orthogonal matrix P that diagonalizes A is

$$P = \frac{1}{2} \begin{bmatrix} \sqrt{2} & -\sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} \\ 1 & 1 & \sqrt{2} \end{bmatrix},$$

and with the change of coordinates $\mathbf{x} = P\mathbf{y}$, that is,

$$x = \frac{1}{\sqrt{2}}(x' - y'), \quad y = \frac{1}{2}(x' + y' - \sqrt{2}z'), \quad z = \frac{1}{2}(x' + y' + \sqrt{2}z'),$$

the equation is transformed to $\sqrt{2}(x')^2 - \sqrt{2}(y')^2 = 1$, which is a hyperbolic cylinder as shown in the figure. Note that $\text{In}(A) = (1, 1, 1)$. $\square$

• **Hyperbolic cylinder:**

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 - y^2 = 1\}.$$

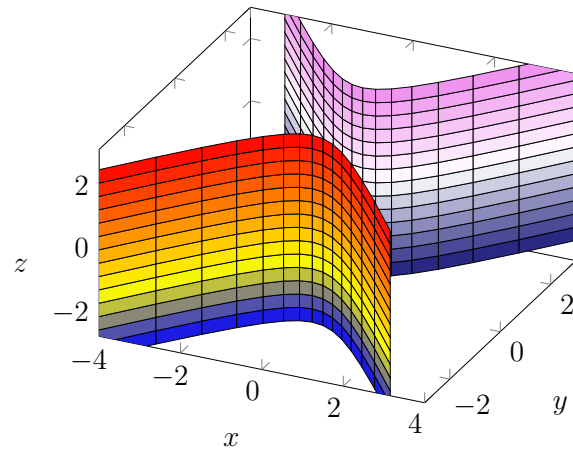


Figure 4.7: Hyperbolic cylinder

• **Elliptical cylinder:**

$$\left\{ (x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \right\}.$$

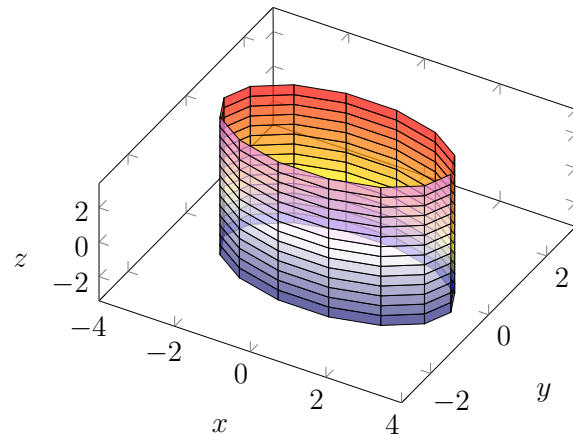


Figure 4.8: Elliptic cylinder

• **Parabolic cylinder:**

$$\{(x, y, z) \in \mathbb{R}^3 : y = x^2\}.$$

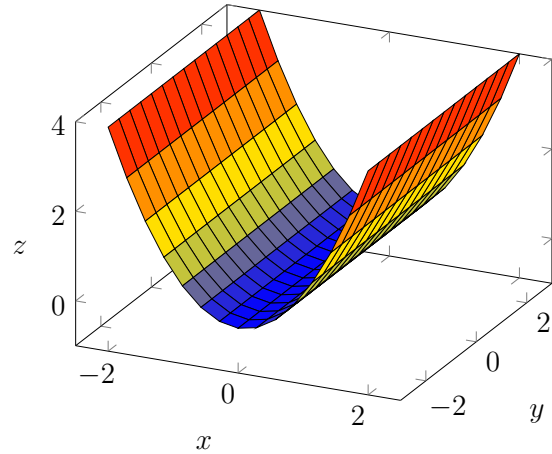


Figure 4.9: Parabolic cylinder

Now, consider a general form of a quadratic equation on $\mathbb{R}^n$

$$\mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} = c.$$

- (1) If it does not have a linear form, i.e., $\mathbf{b} = \mathbf{0}$, then, as shown already, a *parabolic* level surface does not appear as a solution of the quadratic equation.
- (2) Suppose that it has a nonzero linear form, i.e., $\mathbf{b} \neq \mathbf{0}$. If the matrix A is invertible, then, by taking a change of variables as $\mathbf{y} = \mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}$ (or $\mathbf{x} = \mathbf{y} - \frac{1}{2}A^{-1}\mathbf{b}$), (it is a translation) the given quadratic equation is transformed into a new quadratic equation $\mathbf{y}^T A \mathbf{y} = d$ without a linear form, where $d = c + \frac{1}{4}\mathbf{b}^T A^{-1}\mathbf{b}$. However, if A is not invertible, the solution of the quadratic equation depends not only on the inertia of A , but also on the type of linear form, and a parabolic level surface appears as the solution of the quadratic equation with a nonzero linear form. For example, the equation $x^2 - z = c$ has a singular quadratic form for which $\text{In}(A) = (1, 0, 2)$ and also has a nonzero linear form that cannot be removed by any change of variables. The solution of this equation is a **parabolic cylinder** when $n = 3$.

Example 4.7.3 (A quadratic equation having a linear form). Determine the conic section for $3x^2 - 6xy + 4y^2 + 2x - 2y = 0$.

Solution: The matrix for the quadratic form $3x^2 - 6xy + 4y^2$ is

$$A = \begin{bmatrix} 3 & -3 \\ -3 & 4 \end{bmatrix}.$$

Its inverse is

$$A^{-1} = \frac{1}{3} \begin{bmatrix} 4 & 3 \\ 3 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}.$$

With the change of variables $\mathbf{y} = \mathbf{x} + \frac{1}{2}A^{-1}\mathbf{b}$, that is

$$x' = x + \frac{1}{3}, \quad y' = y,$$

the equation is transformed to a new equation $3(x')^2 - 6x'y' + 4(y')^2 = \frac{1}{3}$. Clearly, the matrix representation of the new quadratic form is also A , and its eigenvalues are $\frac{1}{2}(7 \pm \sqrt{37})$. Therefore, $\text{In}(A) = (2, 0, 0)$ and the solution of the equation is an ellipse. $\square$

Exercise 4.7.4. Identify the level surface of each of the following quadratic equations:

- (1) $2x^2 + 2y^2 + 6yz + 10z^2 = 9$;
- (2) $x^2 - 8xy + 16y^2 - 3z^2 = 8$;
- (3) $4x^2 + 12xy + 9y^2 + 3x - 4 = 0$.

We give some more solved examples below.

Example 4.7.5. Identify the surface given by $11x^2 + 6xy + 19y^2 = 80$. Also convert it to the standard form by finding the orthogonal matrix P .

Solution. The symmetric matrix associated with above quadratic form is

$$A = \begin{bmatrix} 11 & 3 \\ 3 & 19 \end{bmatrix}.$$

We find the characteristic roots of A . We know that characteristic roots of A are the solutions of $\det(A - \lambda I) = 0$. Now,

$$\begin{aligned} \det(A - \lambda I) &= 0 \\ \Rightarrow \det \left(\begin{bmatrix} 11 & 3 \\ 3 & 19 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) &= 0 \\ \Rightarrow \begin{vmatrix} 11 - \lambda & 3 \\ 3 & 19 - \lambda \end{vmatrix} &= 0 \\ \Rightarrow (11 - \lambda)(19 - \lambda) - 9 &= 0 \\ \Rightarrow \lambda^2 - 30\lambda + 200 &= 0 \\ \Rightarrow (\lambda - 10)(\lambda - 20) &= 0 \end{aligned}$$

Therefore 10 and 20 are the characteristic roots of A . Let $\begin{bmatrix} x \\ y \end{bmatrix}$ be a characteristic vector of A corresponding to the characteristic root 10. Then

$$\begin{bmatrix} 11 & 3 \\ 3 & 19 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 10 \begin{bmatrix} x \\ y \end{bmatrix}.$$

$$\therefore 11x + 3y = 10x \text{ and } 3x + 19y = 10y.$$

i.e., $x + 3y = 0$. Taking $y = 1$, we have $x = -3$. Therefore, $\frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix}$ is an characteristic vector of A corresponding to the characteristic root 10.

Similarly for $\lambda = 20$,

$$\begin{bmatrix} 11 & 3 \\ 3 & 19 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 20 \begin{bmatrix} x \\ y \end{bmatrix}.$$

$$\therefore 11x + 3y = 20x \text{ and } 3x + 19y = 20y.$$

i.e., $3x - y = 0$. Taking $x = 1$, we have $y = 3$. Hence, $\frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ is an characteristic vector of A corresponding to the characteristic root 20. Therefore, the orthogonal matrix P (i.e., $PP^T = I$) is

$$P = \begin{bmatrix} -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

Now, let

$$\begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}, \text{ i.e.,} \\ x = -\frac{3}{\sqrt{10}}x' + \frac{1}{\sqrt{10}}y' \text{ and } y = \frac{1}{\sqrt{10}}x' - \frac{3}{\sqrt{10}}y'.$$

Substituting these values in the given equation $11x^2 + 6xy + 19y^2 = 80$, we get $10x'^2 + 20y'^2 = 80$ or $x'^2 + 2y'^2 = 8$. Therefore, the standard form of the given quadratic equation is $\boxed{x'^2 + 2y'^2 = 8}$ and it is an *ellipse*.

• **Ellipse:** $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Here, in this example, we have

$$\frac{x'^2}{(2\sqrt{2})^2} + \frac{y'^2}{2^2} = 1.$$

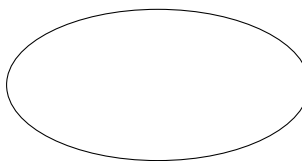


Figure 4.10: An ellipse

□

Example 4.7.6. Reduce the quadratic form $x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$ into standard form by finding an orthogonal matrix P . Hence determine the surface given by $x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3 = 1$.

Solution. The matrix associated with the above quadratic form is $A = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & -1 \end{bmatrix}$. Check that the roots of $\det(A - \lambda I) = 0$ are 3, -3 and 0 and characteristic vectors corresponding to these characteristic roots are $\begin{bmatrix} -2 \\ 3 \\ 3 \\ 1 \\ 3 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 3 \\ 2 \\ 3 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \\ 1 \\ 2 \\ 3 \end{bmatrix}$ respectively.

Hence, the required orthogonal matrix is

$$P = \begin{bmatrix} \frac{-2}{3} & \frac{-1}{3} & \frac{2}{3} \\ \frac{3}{3} & \frac{3}{3} & \frac{3}{3} \\ \frac{3}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{3}{3} & \frac{2}{3} \\ \frac{3}{3} & \frac{3}{3} & \frac{3}{3} \end{bmatrix}.$$

Let $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = P \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, i.e., $x_1 = \frac{-2}{3}y_1 - \frac{1}{3}y_2 + \frac{2}{3}y_3$, $x_2 = \frac{2}{3}y_1 - \frac{2}{3}y_2 + \frac{1}{3}y_3$ and $x_3 = \frac{1}{3}y_1 + \frac{2}{3}y_2 + \frac{2}{3}y_3$. Substituting these values in the quadratic form we get $3y_1^2 - 3y_2^2$. Therefore, the standard form for the quadratic equation is $\boxed{3y_1^2 - 3y_2^2 = 1}$ and it is a *hyperbolic cylinder*. □

Example 4.7.7. Describe the conic C whose equation is $5x^2 - 4xy + 8y^2 + 4\sqrt{5}x - 16\sqrt{5}y + 4 = 0$.

Solution. The matrix associated with the above quadratic equation is $A = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}$. Check that 4 and 9 are the characteristic roots of A and the corresponding characteristic vectors are

$\begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$ and $\begin{bmatrix} \frac{-1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$. Hence, the orthogonal matrix

$$P = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}.$$

Let $\begin{bmatrix} x' \\ y' \end{bmatrix} = P \begin{bmatrix} x \\ y \end{bmatrix}$, i.e., $x' = \frac{2}{\sqrt{5}}x - \frac{1}{\sqrt{5}}y$ and $y' = \frac{1}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y$. Substituting it in the given quadratic form we get $4x'^2 + 9y'^2 + 4\sqrt{5}(\frac{2}{\sqrt{5}}x' - \frac{1}{\sqrt{5}}y') - 16\sqrt{5}(\frac{1}{\sqrt{5}}x' + \frac{2}{\sqrt{5}}y') + 4 = 0$, i.e., $4x'^2 + 9y'^2 - 8x' - 36y' + 4 = 0$. Therefore, $4(x' - 1)^2 + 9(y' - 2)^2 = 36$, which is again an ellipse. $\square$